Optimization Algorithms
Weekly Exercise 4

Marc Toussaint
Learning & Intelligent Systems Lab, TU Berlin
Marchstr. 23, 10587 Berlin, Germany

Winter 2024/25

1 Minimalistic Log Barrier

Consider the 1D function, $x \in \mathbb{R}$,

$$f_\mu(x) = -x - \mu \log(-x)$$

(Note: This is the log barrier function for the problem $\min_{x\in\mathbb{R}} -x$ s.t. $x \le 0$.)

a) Plot the function for varying $\mu = 1, 0.5, 0.1$.

b) Analytically find the mimimum $x^*(\mu) = \text{argmin}_x f_\mu(x)$ as a function of μ.

c) Prove that $\lim_{\mu\to 0} x^*(\mu) = 0$.

2 Dual Update in Augmented Lagrangian

The squared penalty approach to solving an equality constrained optimization problem minimizes in each inner loop:

$$\min_x \; f(x) + \mu \sum_{i=1}^m h_i(x)^2 \; . \tag{1}$$

The Augmented Lagrangian method adds a Lagrangian term and minimizes in each inner loop:

$$\min_x \; f(x) + \mu \sum_{i=1}^m h_i(x)^2 + \sum_{i=1}^m \lambda_i h_i(x) \; . \tag{2}$$

Assume that we first minimize (1) such that we end up at a minimum $\bar{x}$.
Now prove that, under the assumption that the gradients $\nabla f(x)$ and $\nabla h(x)$ are (locally) constant, setting $\lambda_i = 2\mu h_i(\bar{x})$ will ensure that the minimum of (2) fulfills the constraints $h(x) = 0$.

3 Gradient descent with matrices

(This exercise goes beyond the context of the lecture, but further trains you in dealing with derivatives and gradient descent when the decision variable is a matrix.)
One way to derive a gradient is through the Taylor approximation

$$f(w+h) \approx f(w) + \langle \delta, h \rangle$$

where δ is the gradient and $\langle \delta, h \rangle = \delta^T h$ the standard scalar product. Now assume that w is not a vector, but a matrix $W \in \mathbb{R}^{n\times m}$ and let $f(W) = \|WX - Y\|_F^2$ with X, Y matrices of appropriate sizes and $\|\cdot\|_F$ the Frobenius norm. What is the D in

$$f(W+H) \approx f(W) + \langle D, H \rangle_F$$

and how does a gradient step look like?
Tips: $\langle A, B \rangle_F = \text{tr}(A^T B)$